CLAIMS

1. A method of selecting a server to represent a virtual server hosted by a plurality of servers, comprising:

providing, by a load balancer not associated with the virtual server, values, for one or more parameters, of two or more paths, each path defined between a point in a vicinity of a client accessing the virtual server and one of the plurality of servers representing the virtual server; and

selecting a server to provide data for the client, responsive to the values of the one or more parameters.

2. A method according to claim 1, wherein the load balancer and the client are in the same metropolitan area.

3. A method according to claim 1, wherein the load balancer and the client are in the same local area network.

4. A method according to claim 1, wherein the one or more parameters comprise at least one of a jitter, a round trip delay or a hop count.

5. A method according to claim 1, wherein the one or more parameters comprise a cost.

6. A method according to claim 1, wherein selecting the server comprises selecting, by a client-controlled load balancer, responsive to receiving identification of a virtual server requested by the client.

7. A method according to claim 6, wherein selecting the server comprises selecting, by a client-controlled load balancer, responsive to receiving a connection establishment request from the client.

8. A method according to claim 6, wherein providing the values for the one or more parameters comprises measuring at least one of the parameters.

9. A method according to claim 8, wherein measuring at least one of the parameters, for at least one of the paths, is performed before receiving the connection establishment request.

10. A method according to claim 8, wherein measuring at least one of the parameters for at least one of the paths is performed after receiving the connection establishment request.

11. A method according to claim 1, further comprising changing the destination IP address of packets received by the load balancer from the client, to an IP address of the selected server.

12. A method according to claim 1, further comprising changing the source IP address of packets received by the load balancer from the selected server.

13. A method according to claim 1, further comprising transmitting an IP address of the selected server to the client.

14. A method according to claim 13, wherein transmitting the IP address of the selected server to the client comprises transmitting a DNS response.

15. A method according to claim 1, wherein ones of the plurality of servers are located in different geographical regions.

16. A method according to claim 1, wherein selecting a server to provide data for the client comprises selecting, by the load balancer, a second load balancer which is to perform the server selection and selecting, by the second load balancer, a server to provide data for the client.

17. A method according to claim 1, wherein the virtual server hosts a web site.

18. A method according to claim 1, wherein selecting a server to provide data for the client comprises selecting a server which minimizes a function of the one or more parameters.

19. A method according to claim 18, wherein selecting a server to provide data comprises choosing a function of the one or more parameters to be minimized and selecting a server which minimizes the chosen function.

20. A method according to claim 19, wherein the function is chosen responsive to a protocol with which the virtual server is accessed.

21. A method according to claim 19, wherein the function is chosen responsive to the virtual server accessed.

22. A method according to claim 19, wherein the function is chosen responsive to an attribute of the client.

23. A method according to claim 19, wherein the function is chosen responsive to the time of the selection.

24. A method of selecting a server to be accessed, comprising:

receiving, by a load balancer, a message relating to a virtual server, hosted by a plurality of servers, and to a client desiring to receive data from the virtual server; and

selecting, by the load balancer, one of the plurality of servers to provide data to the server,

wherein the load balancer is closer to the client than to the selected server.

25. A method according to claim 24, wherein the load balancer is closer to the client than to any of the plurality of servers hosting the virtual server.

26. A method according to claim 24, wherein the load balancer is in the same metropolitan area as the client.

27. A method according to claim 24, wherein the load balancer is in the same local area network as the client.

28. A method according to claim 24, wherein the load balancer is not associated with the virtual server.

29. A method according to claim 24, wherein the load balancer is under control of a system manager of the client.

30. A method according to claim 24, wherein receiving the message comprises receiving a DNS query message.

31. A method according to claim 24, wherein receiving the message comprises receiving from a DNS server.

32. A method according to claim 24, wherein receiving the message comprises receiving a connection establishment request directed to the virtual server.

33. A method according to claim 24, wherein receiving the message comprises receiving a message directed to the load balancer.

34. A method according to claim 24, wherein selecting one of the servers comprises selecting a server which has a lowest cost path to the load balancer.

35. A method according to claim 24, wherein selecting one of the servers comprises selecting a server which has a lowest delay path or a highest packet size path to the load balancer.

36. A method according to claim 24, wherein the load balancer is geographically closer to the client than to the selected server.

37. A method of selecting a server to be accessed, comprising:

receiving, by a load balancer, a message relating to a virtual server, hosted by a plurality of servers, and to a client desiring to receive data from the virtual server; and

selecting, by the load balancer, one of the plurality of servers to provide data to the client, at least partially responsive to the cost of communications between the client and one or more of the plurality of servers.

38. A method according to claim 37, wherein selecting one of the servers comprises selecting a server under a constraint that a lowest cost client communication connection is used in connecting to the server.

39. A method according to claim 37, wherein selecting one of the servers comprises selecting a server which minimizes a weighted sum of communication costs to the server and at least one other route related parameter.

40. A method according to claim 39, wherein selecting one of the servers comprises selecting a server which minimizes a weighted sum of the communication costs to the server and the round trip delay to the server.

41. A load balancer, comprising:

an interface adapted to receive server access messages from clients; and

a processor adapted to determine, for at least one of the messages, whether the message requires load balancing responsive to at least one attribute different from the identity of the server referenced by the message, and to select for at least one message determined to require load balancing, a server to service the client.

42. A load balancer according to claim 41, wherein the at least one attribute comprises the time at which the message is received at the interface.

43. A load balancer according to claim 41, wherein the at least one attribute comprises the identity of the client.

44. A load balancer according to claim 41, wherein the at least one attribute comprises a protocol to govern the communication with the server.

45. A load balancer according to claim 41, further comprising a packet changing unit adapted to change the contents of at least one field of packets belonging to connections for which load balancing was performed.

46. A load balancer according to claim 41, wherein the packet changing unit is adapted to change packets in accordance with half NAT or full NAT procedures.

47. A method of selecting a server to be accessed, comprising:

receiving, by a load balancer, a message relating to a virtual server, hosted by a plurality of servers, and to a client desiring to receive data from the virtual server;

choosing a function from a plurality of predetermined functions utilized by the load balancer for selecting servers, responsive to the received message; and

selecting, by the load balancer, one of the plurality of servers that minimizes or maximizes the chosen function, to provide data to the client.

48. A method according to claim 47, wherein choosing the function comprises choosing responsive to an identity of the client.

49. A method according to claim 47, wherein choosing the function comprises choosing responsive to a time at which the message is received.

50. A method according to claim 47, wherein at least two of the predetermined functions depend on different groups of one or more parameters.

51. A method according to claim 47, wherein at least two of the predetermined functions depend on the same parameters but give different weight to one or more of the parameters on which they depend.